Cloud platform for data management of the environmental monitoring network: UNECE ICP vegetation case

M. Frontasyeva, N. Kutovskiy, A. Nechaevskiy, G. Ososkov, A. Uzhinskiy[a]

Joint Institute for Nuclear Research, 6 Joliot-Curie, 141980, Dubna, Russia

E-mail: [a]uzinskiy@jinr.ru

The aim of the UNECE ICP Vegetation program is to identify the main polluted areas, produce regional maps and further develop the understanding of the long-range transboundary pollution. Since January 2014, the coordination of moss surveys in 36 European and Asian countries has been conducted from the JINR in Russia. To manage monitoring data a cloud platform is proposed. It consists of a set of inter-connected services and tools deployed and hosted in the JINR cloud. Motivation, basic principles and architecture of the platform are presented.

Keywords: environmental monitoring, data management, cloud platform, services, analysis, mosses.

© 2016 M. Frontasyeva, N. Kutovskiy, A. Nechaevskiy, G. Ososkov, A. Uzhinskiy,

1 Introduction

Air pollution has a significant negative impact on the various components of ecosystems, human health, and ultimately, cause significant economic damage. Increased ratification of the Protocols of the Convention on Long-range Transboundary Air Pollution (LRTAP) is identified as a high priority in the new long-term strategy of the Convention. Full implementation of air pollution abatement policies is particularly desirable for countries of Eastern Europe, the Caucasus and Central Asia (EECCA) as well as South-Eastern Europe (SEE). Atmospheric deposition study of heavy metals, nitrogen, persistent organic compounds and radionuclides is based on the analysis of naturally growing mosses through moss surveys carried out every 5 years [UNECE ICP Vegetation]. Due to intense activity of the Joint Institute for Nuclear Research (JINR) as a coordinator of the moss surveys since 2014, Azerbaijan, Belarus, Georgia, Kazakhstan, Moldova, Turkey and Ukraine participated in the moss survey for 2015/2016. Nowadays the UNECE ICP Vegetation programme [Harmens and Mills, 2014] is realized in 36 countries of Europe and Asia. Mosses are collected at thousands of sites across Europe and their heavy metal (since 1990), nitrogen (since 2005), POPs (Persistent organic compounds, pilot study in 2010) and radionuclides (since 2015) concentrations are determined. The goal of this study program is to identify the main polluted areas, produce regional maps and further develop the understanding of long-range transboundary pollution [Harmens et al, 2015].

2 Experiment and data interpretation

Sampling is carried out in compliance with the internationally accepted guidelines [EUROPEAN MOSSES: 2015 SURVEY]. Such analytical techniques as AAS, AFS, CVAAS, CVAFS, ETAAS, FAAS, GFAAS, ICP-ES, ICP-MS, as well as INAA are used for elemental determination. A total of 13 elements are reported to the Atlas (As, Cd, Cr, Cu, Fe, Hg, Ni, Pb, V, Zn, Al, Sb, and N). Nowadays POPs (whichever determined) and radionuclides (namely, 210Pb and 137Cs) are accepted for air pollution characterization. The results are reported as number of sampling sites, minimum, maximum and median concentrations in mg/kg. The data interpretation is based on Multivariate statistical analysis (factor analysis), description of sampling sites (MossMet information package) and distribution maps for each element produced using ArcMAP, part of ArcGIS, an integrated geographical information system (GIS) [Buse et al, 2003]. Examples of GIS maps are presented on Fig. 1.

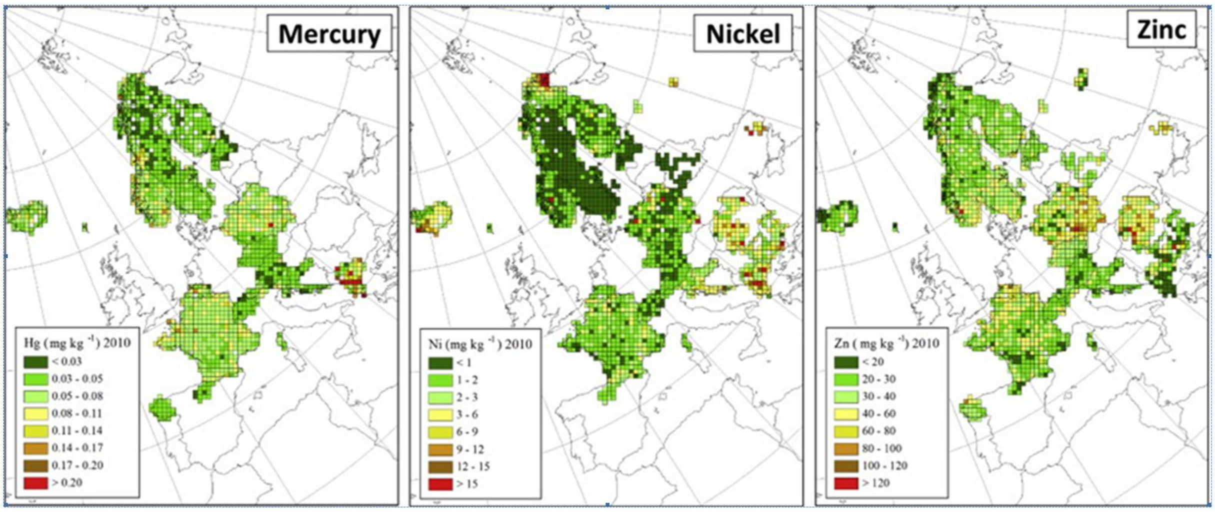

Fig. 1. Examples of distribution maps [3]

3 Motivation

The ICP Vegetation programme is very important project, but it has a serious weakness related to its weak adoption of modern informational technologies. There are dozens of respondents in the existing monitoring network and their number is increasing, but information on collecting and processing of samples is carried out manually or with minimum automation. Data mostly stored in xls files and aggregated manually by the coordinator. Files from respondents are usually passed to the coordinator by email or by ordinary mail. There are no common standards in data transfer, storing and processing software. Such situation does not meet the modern standards for quality, effectiveness and speed of research. Lack of a single web-platform that provides comprehensive solution of biological monitoring and forecasting tasks is a major problem for research.

Therefore, the aim of the project is to create a cloud platform using modern analytical, statistical, programmatic and organizational methods to provide the scientific community with unified system of gathering, storing, analyzing, processing, sharing and collective usage of biological monitoring data.

The platform elements are to facilitate IT-aspects of all biological monitoring stages starting from a choice of collection places and parameters of samples description and finishing with generation of pollution maps of a particular area or state-of-environment forecast in the long term. Mechanisms and tools for association of participants of heterogeneous networks of biological monitoring are provided in the platform. That enables verifying obtained results and optimizes research. The open part of the platform can be used for informing public authorities, local governments, legal entities and individuals about state-of-environment changes.

One more important aspect of ecological researches relates to various statistical methods applied to process collected data. Modern approaches to explore air pollutions provided by heavy metals, nitrogen, POPs and radionuclides include as a mandatory part of multivariable statistical and intellectual data processing. Latest tendencies in that field include extension of a set of georeferenced data integrated in data processing of surveyed material. So it is not limited by geographical, topographical or geological information, what is traditional in such cases, but also includes, for example, satellite imagery and their products, topographic high-precision data derived from aerial photography, etc. These new data classes, contrary to the traditional ones, are characterized by a high resolution and dynamic nature – for example, satellite images represents a reflection of solar radiation, which depends on the time of day, season, cloud cover, etc. This in turn greatly increases the amount of data to be processed. The task of integration of different data types is tied to the problem of the development of new models and algorithms – such as neural networks [Alijagić, 2013], self-organizing maps [Žibret and Šajn, 2010], etc. – during the study of dynamic properties of ecological processes among other things.

So, one more aim of our project is to develop modern software tools for multivariable statistical and intellectual data processing oriented on the GIS-technology.

4 Architecture and technologies

To optimize the whole procedure of data management, it is proposed to build a unified platform consisting of a set of interconnected services and tools to be developed, deployed and hosted in the JINR cloud [Baranov et al, 2016]. Such an approach also allows scaling cloud resources up and down according to the service load. When some component will require more resources the cloud can provide them without affecting other components. This increases the efficiency of hardware utilization as well as the reliability and availability of the service itself for end-users. Such auto-scaling behavior will be achieved by using the OneFlow component of the OpenNebula platform, which the JINR cloud is based.

We defined requirements for the platform and it components. The general architecture of the platform and technologies used are depicted in Fig. 2.

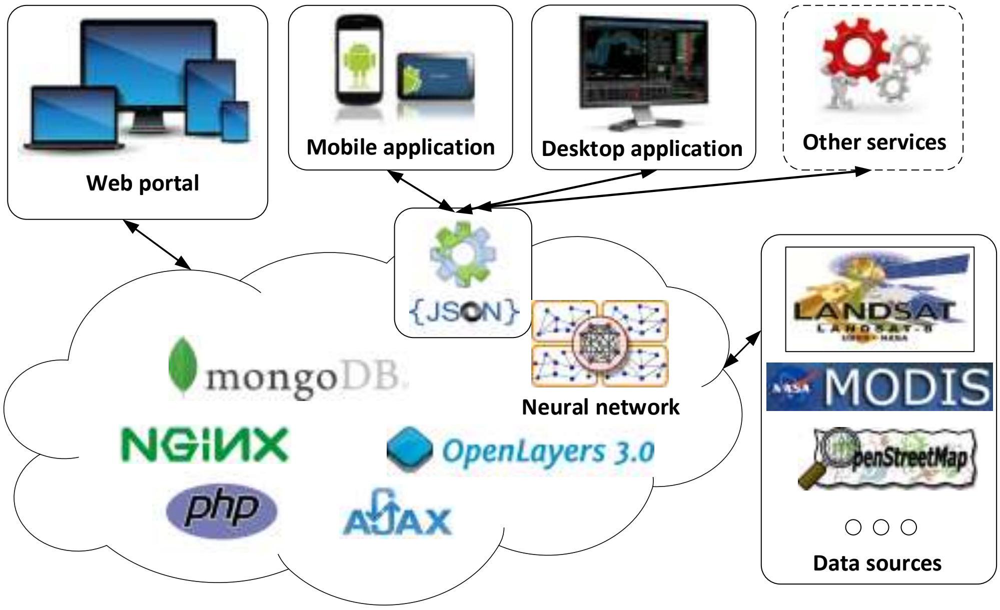

Fig. 2. General architecture of the platform and technologies used

A preliminary data analysis of raw data coming from the contributors are performed. The data samples can have 10 to 40 metrics depending on the collecting area. Most of the metrics are optional, so traditional relation databases will be ineffective. There is also a need to have a possibility to change structure of the data sample object without hard code modification in order to integrate new projects and experiments into the platform easily. Our team has a positive experience with MongoDB (open-source, document database designed for ease of development and scaling) at our previous projects where more than 5 million data records from 200+ contributors are processed so it was decided to use the data base to store sampling results.

The portal back-end built on Nginx (an open source reverse proxy web server for HTTP protocol) and developed with PHP (widely-used open source general-purpose scripting language that is especially suited for web development). That should provide necessary performance and scalability. Calculation of the factors, indexes and other statistic parameters is done by PHP.

The web-portal with responsive design that adjusts to different screen sizes is the main interface of the platform. The portal allows multilevel access to the data and has advanced data processing and reporting mechanisms. There are two parts of the portal – public and private ones. A general information about the project and the platform is presented in the public part. The private part can be accessed only by authorized contributors. Users can manage the project/regions and data samples in the private part. Data samples can be added manually or imported from xls files. Users can get statistical information about their datasets and build distribution or graduated symbol maps. Some screenshots of web-portal interface are shown on the Fig. 3

QGIS (Open Source Geographic Information System) and OpenLayers (opensource javascript library to load, display and render maps from multiple sources) was evaluated by us for regional and global maps representation in the portal. But QGIS and its web plugin is too hard to maintain and develop. So the OpenLayers and some of its specific layers are used to create maps in the portal.

Apart from the graphical user web-interface to the platform it is planned to provide a service with RESTful API for the mobile and desktop applications as well as for third-party services whose users can be interested in the environmental monitoring data.

Data import and export mechanism is available in the platform. Such way users can process data online or upload it and use their local processing application.

Prediction is an important step of the data analysis of any ecological survey. An application of prediction methods enables a mapping of estimate values. Maps in their turn provide visualization of spatial variability of data and can be used for visual analysis so that ecological hazards can be identified.

A neural network together with satellite spectroradiometer products such as surface reflectance, land surface temperature, land cover, vegetation indices, land use are going to be applied to model the statistical relations between element concentrations in moss and potential explanatory variables. This approach has been already tried on the modeled data. A result was a nice concentration map with quite law discrepancy.

LANDSAT, MODIS and OpenSteetMap are considered as source of the satellite and GIS data. Currently the work on import of the satellite spectroradiometer and other useful for neural network data is in progress. Another actual task is determination of the correlated parameters for concentration distribution and integration of the neural network program and the portal.

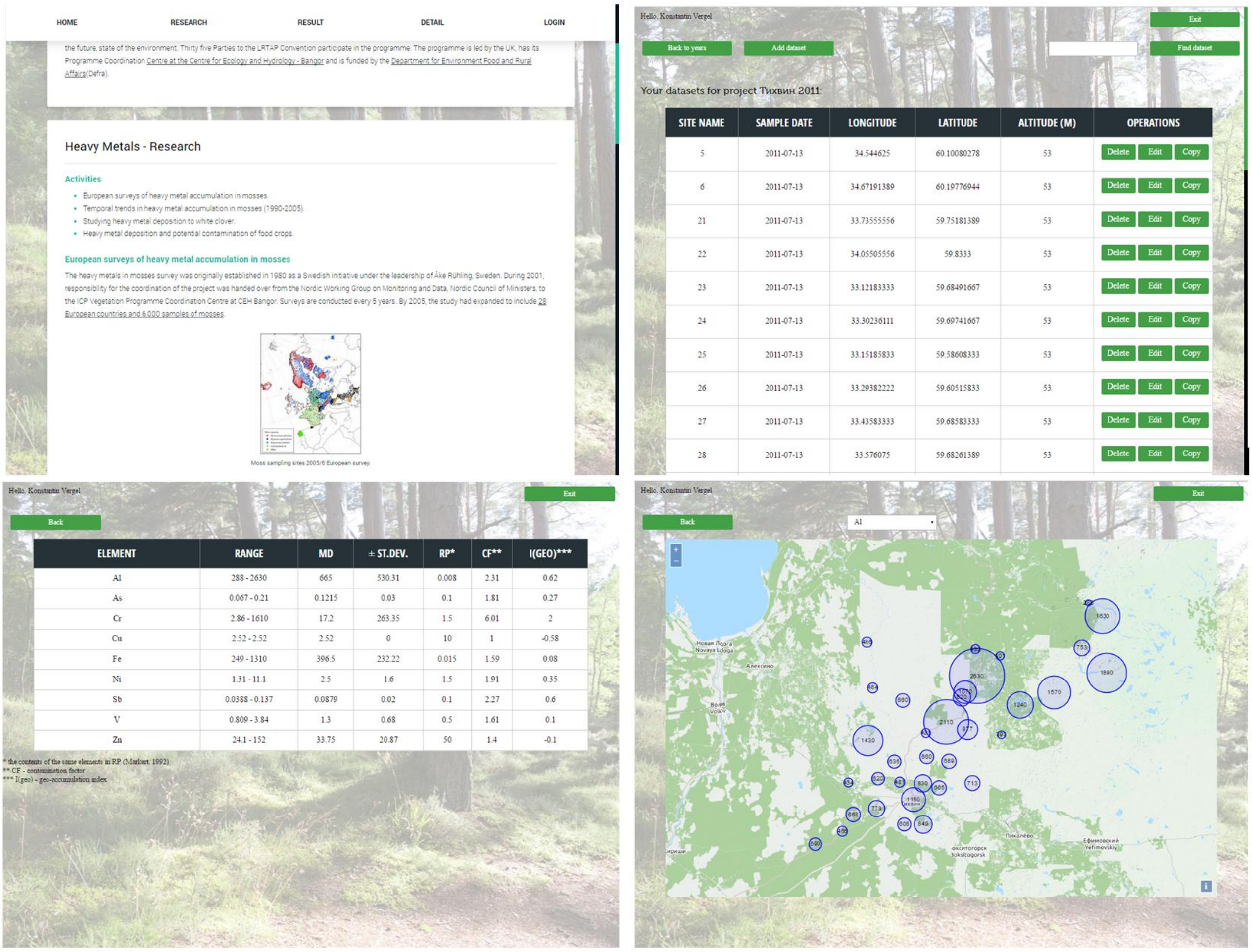

Fig. 3. Web-portal interfaces. Top-left – public part of the portal. Top-right – dataset management interface. Bottom-left – Statistics, factors and indexes calculation for dataset. Bottom-right – graduated symbol map for dataset

5 Conclusion

The study of migration and accumulation of highly toxic pollutants, which include heavy metals, persistent organic pollutants and radionuclides, the influence of pollutants on the various components of the natural and urban ecosystems is the key problem of modern biogeochemistry and ecology. The aim of the given project is to create cloud platform using modern analytical, statistical, programmatic and organizational methods to supply the scientific community with unified system of collecting, analyzing and processing of biological monitoring data. The project is carried out in the framework of the International Cooperative Program ICP Vegetation.

References

United Nations Economic Commission for Europe International Cooperative Programme on Effects of Air Pollution on Natural Vegetation and Crops [Electronic resource]: http://icpvegetation.ceh.ac.uk/

Harmens H. and Mills G. (Eds.) Air Pollution: Deposition to and impacts on vegetation in (South-East Europe, Caucasus, Central Asia (EECCA/SEE) and South-East Asia. Report prepared by ICP Vegetation, March 2014. ICP Vegetation Programme Coordination Centre, Centre for Ecology and Hydrology, Bangor. UK. — 2014. — 72p.

Harmens H., Norris D.A., Sharps K., Mills G., Frontasyeva M., et al. Heavy metal and nitrogen concentrations in mosses are declining across Europe whilst some “hotspots” remain in 2010. // Environmental Pollution. — 2015. — p. 93-104.

HEAVY METALS, NITROGEN AND POPs IN EUROPEAN MOSSES: 2015 SURVEY: http://icpvegetation.ceh.ac.uk/publications/documents/MossmonitoringMANUAL-2015-17.07.14.pdf

Buse A. et al. Heavy metals in European mosses: 2000/2001 survey. UNECE ICP Vegetation Coordination Centre, Centre for Ecology and Hydrology, Bangor, UK. — 2003.

J. Alijagić. Application of multivariate statistical methods and artificial neural network for separation natural background and influence of mining and metallurgy activities on distribution of chemical elements in the Stavnja valley (Bosnia and Herzegovina) // PhD thesis. University of Nova Gorica. — 2013.

Žibret, G., Šajn, R. Hunting for Geochemical Associations of Elements: Factor Analysis and Self-Organising Maps. // Mathematical Geosciences. — 2010 — 42(6): 681–703.

A.V. Baranov, N.A. Balashov, N.A. Kutovskiy, R.N. Semenov JINR cloud infrastructure evolution // Physics of Particles and Nuclei Letters. . — 2016 — ISSN 1547-4771 — eISSN: 1531-8567 . — vol. 13 — No. 5 — pp. 672–675 — DOI: 10.1134/S1547477116050071.